# Model Archive Summary for Suspended-Sediment Concentration at U.S. Geological Survey Station 11455350; Cache Slough at Ryer Island, California

This model archive summary describes the suspended-sediment concentration (SSC) model developed to compute a 15-minute SSC time-series for the period of record: July 16, 2008 to July 9, 2013. This is the first suspended-sediment model developed for the site. The methods used follow U.S. Geological Survey (USGS) guidance as referenced in the Office of Surface Water/Office of Water Quality Technical Memorandum and USGS Techniques and Methods, book 3 chapter 4 (USGS, 2016; Rasmussen and others, 2009). This summary and model archive are in accordance with Attachment A of Office of Water Quality Technical Memorandum 2015.01 (USGS, 2014).

## Site and Model Information

Site number: 11455350
Site name: Cache Slough at Ryer Island, California (RYI)
Location: Latitude 38°12'46", Longitude 121°40'09" referenced to North American Datum of 1983, Solano County, CA, Hydrologic Unit 18020163.
Equipment: A YSI 6-series sonde began logging turbidity with a model 6136 sensor on July 16, 2008 and was removed on July 9, 2013.

Model number: 11455350.SSC.WY08.1
Model calibration data period: August 15, 2008 – June 26, 2013
Model application date: July 16, 2008 – July 9, 2013
Computed by: Anna Conlen, USGS, Sacramento, CA (aconlen@usgs.gov)
Reviewed by: Tara Morgan-King, USGS, Sacramento, CA (tamorgan@usgs.gov)

## Physical Sampling Details and Sediment Data

All sediment data were collected using USGS protocols and are stored in the National Water Information System (NWIS) database: https://waterdata.usgs.gov/nwis (USGS, 2006). Discrete, boat-based samples were collected seasonally, spanning the range of site conditions and specifically targeting large sediment transport events.

Sample collection is consistent with approved field methods described in Edwards and Glysson (1999) and USGS (2006). The Equal Discharge Increment (EDI) method was used to determine the locations of five sampling verticals along the transect where discharge-weighted, suspended sediment samples were collected. The EDI method was used because velocities are not always isokinetic due to the tidal nature of the site (from Table 4-5 of TWRI09A4; USGS, 2006). A boat-based discharge measurement was collected immediately before EDI sampling with an Acoustic Doppler Current Profiler (ADCP) to determine the location of each vertical. A Federal Interagency Sedimentation Project (FISP) US D-96 bag sampler was used to collect depth-integrated samples. The channel cross section can approach 56 feet deep in the thalweg with a mean depth of approximately 37 feet. Velocities during the model calibration data period ranged from -2.81 ft/s to +4.01 ft/s. Sediment at this station is mostly fines (97% fines on average) and any potential sampling bias due to non-isokinetic sampling is considered minimal.

Samples collected before January 2012 were analyzed for SSC (mg/L) by the filtration method at the USGS Sediment Laboratory in Marina, California, while those collected after January 2012 were analyzed for SSC by the USGS Sediment Laboratory at its current location in Santa Cruz, California. Many samples

were also analyzed for the percentage of fines (<0.063 mm), which can be used to identify outliers. Each of the five EDI verticals were analyzed individually by the lab for quality control purposes. The average SSC from these five verticals was computed and used in the calibration dataset. Sediment results are publicly available on NWIS.

All sediment data were reviewed and approved in the USGS NWIS Water-Quality System database (QWDATA) before being applied in the calibration model.

## Surrogate Data

Continuous, 15-minute turbidity data, reported in Formazin Nephelometric Turbidity Units (FNU) and hourly, tidally-filtered discharge data (QFT), reported in cubic feet per second (cfs), were evaluated as explanatory variables for SSC. Turbidity and QFT time-series data were collected by the USGS California Water Science Center and are located at: https://waterdata.usgs.gov/ca/nwis/uv/?site_no=11455350. Turbidity data were analyzed and approved per USGS guidelines (Wagner and others, 2006). Surrogate QFT data used in the calibration model were computed, reviewed and approved before using in the sediment calibration model. Methods to compute discharge (and thus tidally-filtered discharge) follow Levesque and Oberg (2012).

## Model Calibration Dataset

The USGS Surrogate Analysis and Index Developer Tool (SAID) was used to pair surrogate data with discrete sediment data (Domanski and others, 2015). Turbidity and QFT values were paired with each of the 34 sediment samples with a matching window of ± 15 minutes and ± 30 minutes, respectively. The SAID manual is available at: https://pubs.er.usgs.gov/publication/ofr20151177.

Two samples (November 7, 2008 and May 15, 2012) did not have associated turbidity values due to deletions in the time series, and therefore could not be included in the calibration model dataset.

Two EDI samples were collected at the site on August 15, 2008, October 8, 2008, May 27, 2009 and December 6, 2012. The sampling time span for the two samples on each date exceeds one hour and the times of the two sample averages are over 45 minutes apart. Both samples were included in the calibration dataset for each date.

The final calibration dataset is compiled from 32 concurrent measurements of SSC, turbidity and QFT. Summary statistics and the complete model calibration dataset are provided in the following sections.

## Model Development

Simple linear regression (SLR) models and multiple linear regression (MLR) models were assessed using methods described in Helsel and Hirsch (2002). Four models were evaluated: Model 1) linear model with one explanatory variable (turbidity), Model 2) $\log_{10}$-transformed model with one explanatory variable (turbidity), Model 3) linear model with two explanatory variables (turbidity and QFT) and Model 4) $\log_{10}$-transformed model with two explanatory variables (turbidity and QFT).

Diagnostic statistics and plots for model review were output using a combination of Matlab, SAID and the R environment (R Core Team, 2018). Table 3 in Rasmussen and others (2009) shows the best statistical diagnostics to help evaluate regression models. The best model was chosen based on residual plots, coefficient of determination ($R^2$), root-mean-squared error (RMSE), mean square prediction error

(MSPE), significance tests (p-values) and prediction error sum of squares (PRESS) statistics. RMSE and PRESS statistics cannot be used to compare regressions with different response variable units, so $R^2$, MSPE values and residual plots were used as the main determinants of model strength when comparing $log_{10}$-transformed and untransformed models. Values for these statistics were computed for four models and are included in the table below. The best SLR model is a log model with $log_{10}$-transformed turbidity as the surrogate (highlighted in table below). Though the linear model has a higher $R^2$ value and lower MSPE value than the $log_{10}$-transformed model, the residual plot and normal probability of residuals plot indicate the log model best fits the data. The normal probability plot of residuals for the log model is approximately linear with less skew than the linear model, and the plot of residual vs fitted values for the log model suggests that the variances of the residuals are approximately equal.

QFT was not considered further as an explanatory variable because: 1) QFT was not significant in either MLR model (p-value > 0.05), 2) the MLR model calibration datasets contain either 31 or 28 observations, though a total of 48 samples is recommended when a second explanatory variable is included (USGS, 2016) and 3) including QFT in the final model would limit the computed time-series to an hourly record rather than a 15-minute record.

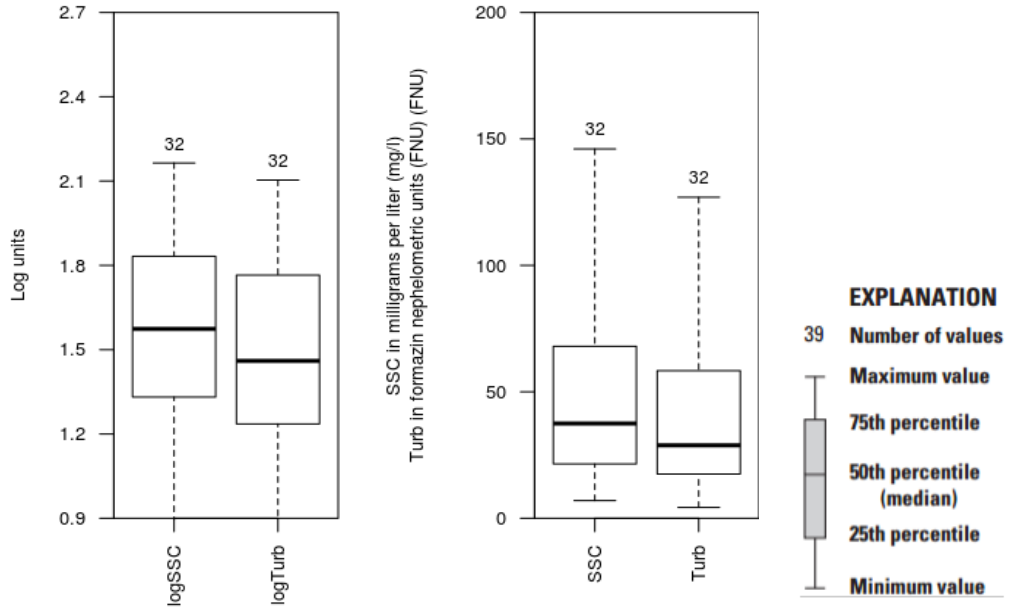
| No. | $R^2$ | $R^2_a$ | RMSE | PRESS | MSPE | n | Type |
|---|---|---|---|---|---|---|---|
| Model 1 | 0.937 | 0.935 | 8.421 | 2409 | 17.8 | 32 | Linear |
| Model 2 | 0.931 | 0.929 | 0.088 | 0.28 | 20.50 | 32 | Log |
| Model 3 | 0.939 | 0.935 | 8.440 | 2355 | 17.5 | 31 | Multi-linear |
| Model 4 | 0.920 | 0.914 | 0.09 | 0.25 | 19.7 | 28 | Multi-Log |

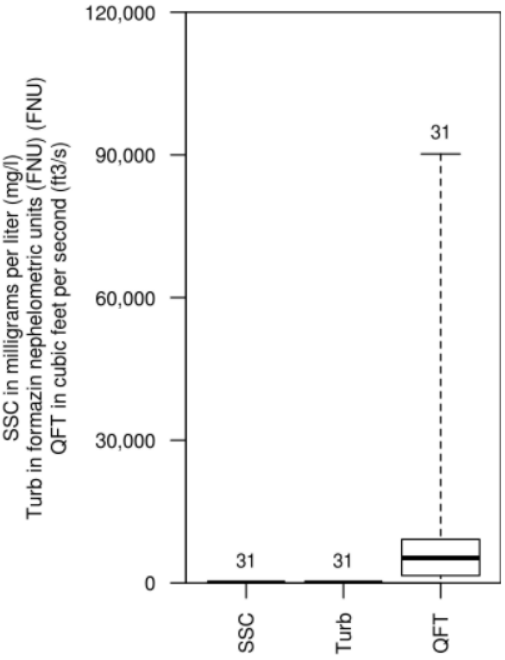
Flagged observations from the SAID outlier test criteria were evaluated. Studentized residuals from the models were inspected for values greater than three or less than negative three; values outside this range are considered potential extreme outliers. The studentized residuals were reviewed from the output reports and none of the samples were deemed to be extreme outliers. All 32 observations were retained in the model.

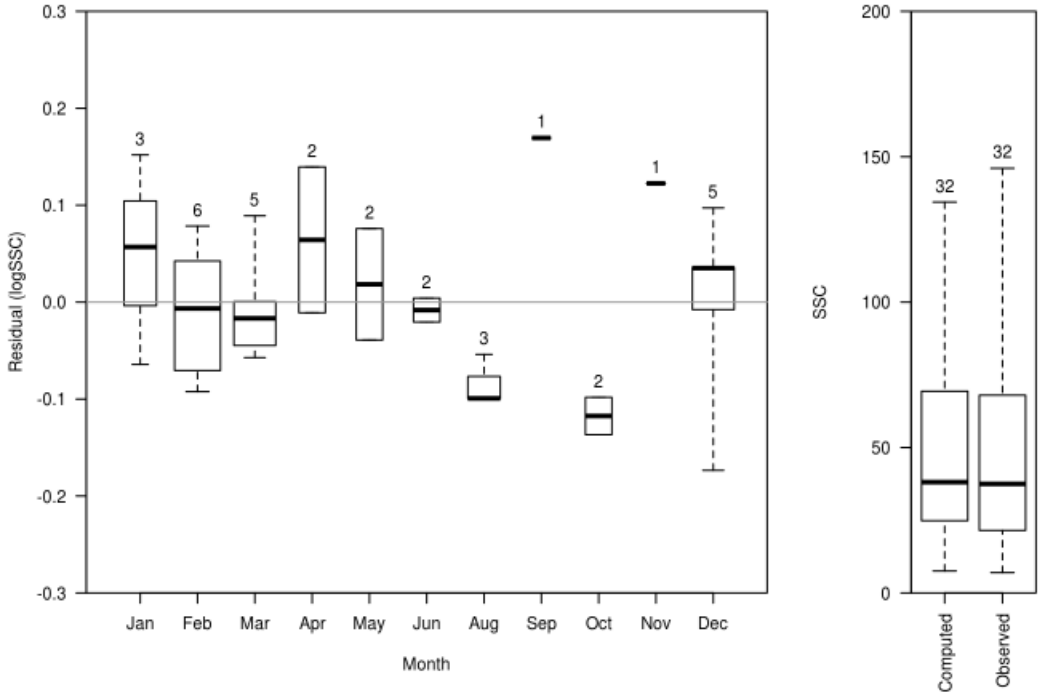
## Plots

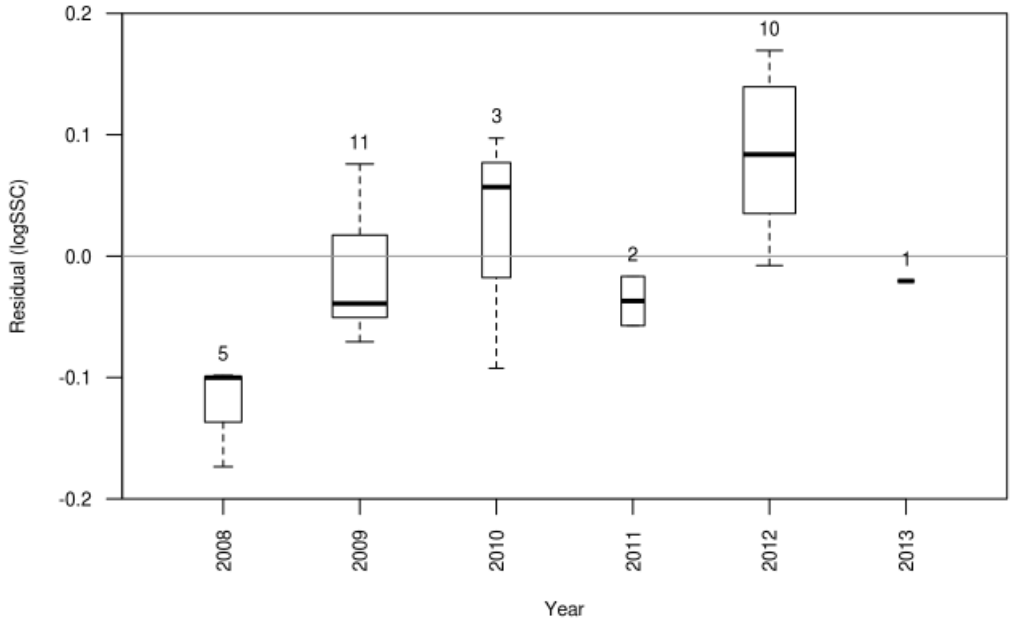
The following plots were generated using a R-based application (Version 1.0) developed by Patrick Eslick of the USGS Kansas Water Science Center. It is available at: http://kswsc.cr.usgs.gov:3838/peslick/ModelArchiveSummary/.

Boxplots of turbidity, QFT and SSC data show the range of measured data for each parameter. The third set of boxplots show SSC residuals of the SLR model by month and water year.


Log units
2.7
2.4
2.1
1.8
1.5
1.2
0.9
32
32
logSSC
logTurb
SSC in milligrams per liter (mg/l)
Turb in formazin nephelometric units (FNU) (FNU)
200
150
100
50
0
32
32
SSC
Turb
EXPLANATION
39 Number of values
Maximum value
75th percentile
50th percentile (median)
25th percentile
Minimum value

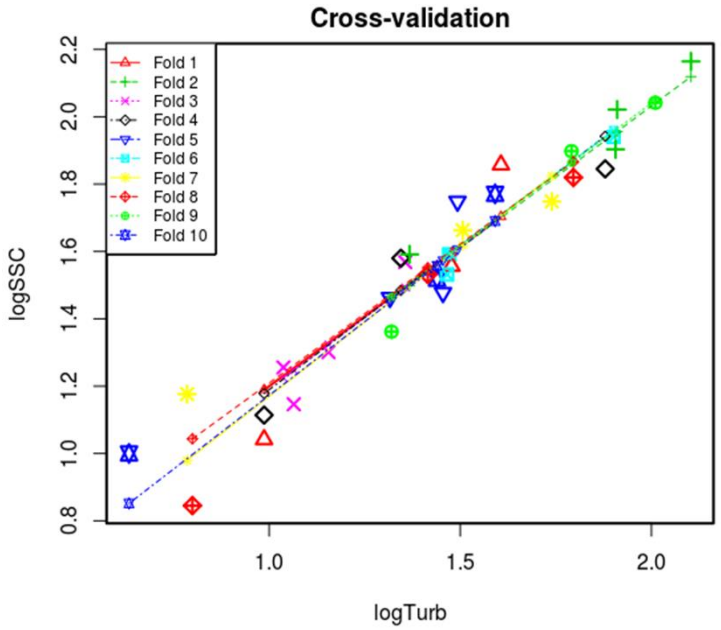

SSC in milligrams per liter (mg/l)
Turb in formazin nephelometric units (FNU) (FNU)
QFT in cubic feet per second (ft3/s)
120,000
90,000
60,000
30,000
0
31
31
31
SSC
Turb
QFT


Residual (logSSC)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
3
6
5
2
2
2
3
1
2
1
5
Jan
Feb
Mar
Apr
May
Jun
Aug
Sep
Oct
Nov
Dec
Month
SSC
200
150
100
50
0
32
32
Computed
Observed

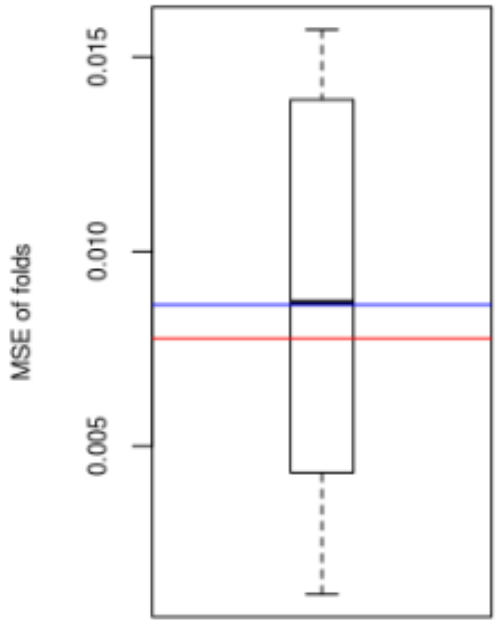

Residual (logSSC)
0.2
0.1
0.0
-0.1
-0.2
5
11
3
2
10
1
2008
2009
2010
2011
2012
2013
Year

## Cross Validation

The cross-validation plot below shows a k-fold validation with k=10 for the final model. The points represent observations that were left out of each fold.




```
                Minimum MSE of folds:  0.00120
                   Mean MSE of folds:  0.00863
                 Median MSE of folds:  0.00868
                Maximum MSE of folds:  0.01570
(Mean MSE of folds) / (Model MSE):  1.11000
```




Red line - Model MSE
Blue line - Mean MSE of folds

## Model Summary

The final SSC model at RYI is a $\log_{10}$-transformed SLR model based on 32 concurrent measurements of SSC and turbidity collected over approximately six water years. The model is shown below with basic model information, regression coefficients, correlation, summary statistics and Duan's bias correction factor (Duan, 1983).

| Linear Regression Model | Coefficient of Determination ($R^2$) |
|---|---|
| $log_{10}SSC = 0.338 + 0.851 * log_{10}Turb$ | 0.931 |

where
SSC = suspended-sediment concentration, in milligrams per liter (mg/L) and
Turb = turbidity, in formazin nephelometric units

SSC was transformed during regression model development, so the computed prediction may be biased and needs to be multiplied by a non-parametric smearing bias correction factor (BCF) when it is retransformed, shown below.

| Model | Start date | End date | Linear Regression Model | *BCF* |
|---|---|---|---|---|
| 1 | 07/16/2008 | 07/09/2013 | $SSC = 10^{0.338} \times Turb^{0.851} \times BCF$ | 1.02 |

The SSC time-series is computed from USGS turbidity data. Minimum and maximum turbidity values of the model application period are listed below. SSC time-series data exceeding extrapolation limits were removed. This model cannot be used to extrapolate more than 10% above or below the range of samples in the calibration dataset (USGS, 2016). The extrapolated, maximum computed SSC for this model is 161 mg/L. The original maximum, computed SSC was 227 mg/L.

| Parameter | Minimum | Maximum |
|---|---|---|
| Computed SSC (mg/L) | 0.8 | 161 |
| Turbidity (FNU) | 0.3 | 229 |

## Suspended-Sediment Concentration Record

The SSC record is computed using this regression model on the USGS National Real-Time Water Quality (NRTWQ) website. The complete record can be found at: https://nrtwq.usgs.gov/ca.

**Model**

log10SSC = 0.338 + 0.851log10Turb

**Variable Summary Statistics**

| | Turb | log10Turb | SSC | log10SSC |
|---|---|---|---|---|
| Minimum | 4.30 | 0.63 | 7 | 0.85 |
| 1st Quartile | 17.50 | 1.24 | 21.50 | 1.33 |
| Median | 28.85 | 1.46 | 37.50 | 1.57 |
| Mean | 38.18 | 1.44 | 47.25 | 1.56 |
| 3rd Quartile | 58.35 | 1.77 | 68 | 1.83 |
| Maximum | 127 | 2.10 | 146 | 2.16 |

**Basic Model Statistics**

| | |
|---|---|
| Number of observations | 32 |
| Root Mean Squared Error (RMSE) | 0.09 |
| Model Standard Percentage Error (MSPE) | 20.5 |
| Coefficient of determination ($R^2$) | 0.931 |
| Adjusted $R^2$ | 0.929 |
| Bias Correction Factor | 1.020 |

**Explanatory Variables**

| | Coefficients | Standard Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 0.338 | 0.06 | 5.38 | 7.91E-06 |
| log10Turb | 0.851 | 0.04 | 20.13 | 5.64E-19 |

**Correlation Matrix**

| | Intercept | E.vars |
|---|---|---|
| Intercept | 1.000 | -0.969 |
| E.vars | -0.969 | 1.000 |

**Outlier Test Criteria**

| Leverage | Cook's D | DFFITS |
|---|---|---|
| 0.188 | 0.193 | 0.500 |

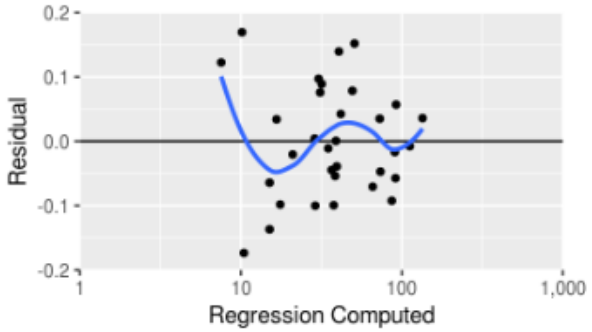
**Flagged Observations**

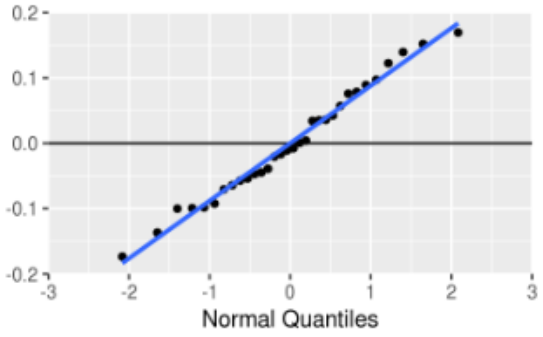
| Date | Time | LogSSC | Estimate | Residual | Standard Residual | Studentized Residual | Leverage | Cook's D | DFFITS |
|---|---|---|---|---|---|---|---|---|---|
| 12/2/2008 | 12:03 | 0.845 | 1.02 | -0.173 | -2.110 | -2.240 | 0.126 | 0.319 | -0.851 |
| 9/25/2012 | 12:12 | 1.180 | 1.01 | 0.169 | 2.060 | 2.190 | 0.130 | 0.318 | 0.846 |
| 11/8/2012 | 10:53 | 1.000 | 0.878 | 0.122 | 1.540 | 1.570 | 0.181 | 0.261 | 0.740 |

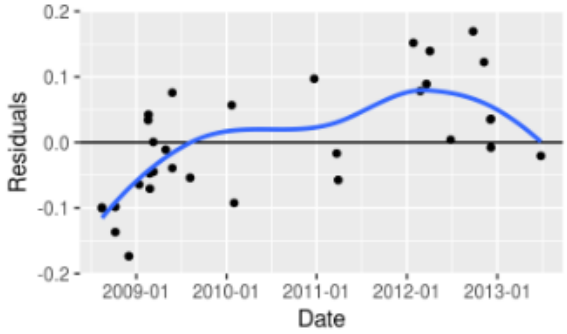
## Residual diagnostic plots

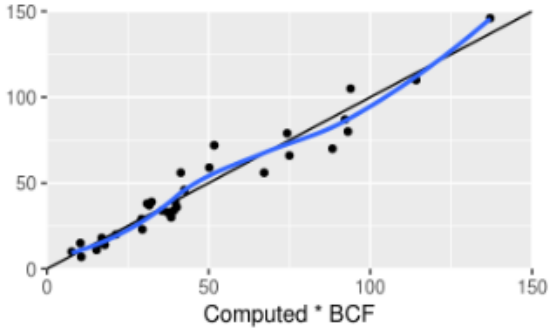
Plots were generated using the model archive summary application developed by Patrick Eslick of the USGS Kansas Water Science Center.

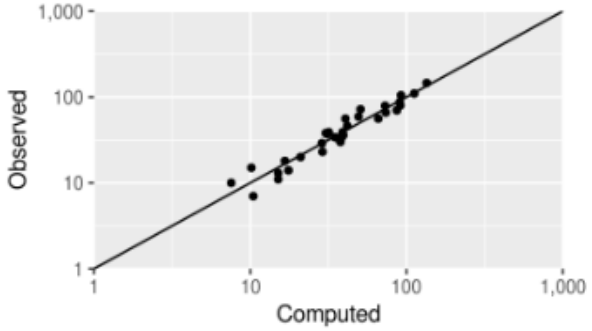
## Statistical Plots

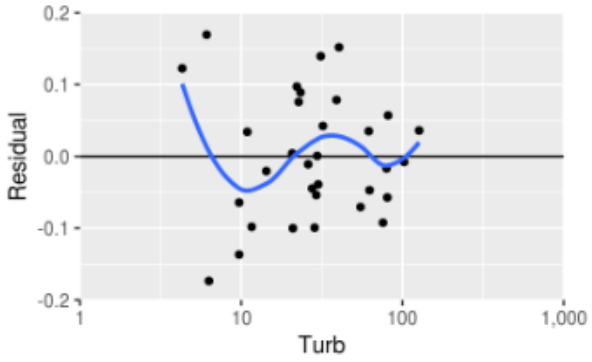


## Model-Calibration Dataset

| | Date & Time | logSSC | logTurb | SSC | Turb | Computed logSSC | Computed SSC | Residual | Normal Quantiles | Censored Values |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | | | |
| 1 | 8/15/2008 12:39 | 1.48 | 1.45 | 30 | 28.5 | 1.58 | 38.4 | -0.0993 | -1.22 | -- |
| 2 | 8/15/2008 14:13 | 1.36 | 1.32 | 23 | 20.9 | 1.46 | 29.5 | -0.1 | -1.4 | -- |
| 3 | 10/8/2008 11:11 | 1.15 | 1.06 | 14 | 11.6 | 1.24 | 17.9 | -0.0982 | -1.07 | -- |
| 4 | 10/8/2008 13:02 | 1.04 | 0.987 | 11 | 9.7 | 1.18 | 15.4 | -0.137 | -1.65 | -- |
| 5 | 12/2/2008 12:03 | 0.845 | 0.799 | 7 | 6.3 | 1.02 | 10.6 | -0.173 | -2.08 | -- |
| 6 | 1/13/2009 12:05 | 1.11 | 0.987 | 13 | 9.7 | 1.18 | 15.4 | -0.0643 | -0.719 | -- |
| 7 | 2/18/2009 16:01 | 1.26 | 1.04 | 18 | 10.9 | 1.22 | 17 | 0.0341 | 0.276 | -- |
| 8 | 2/19/2009 9:40 | 1.66 | 1.51 | 46 | 32.1 | 1.62 | 42.5 | 0.0424 | 0.53 | -- |
| 9 | 2/24/2009 13:24 | 1.82 | 1.8 | 66 | 62.5 | 1.87 | 75 | -0.0472 | -0.442 | -- |
| 10 | 2/25/2009 13:12 | 1.75 | 1.74 | 56 | 54.9 | 1.82 | 67.2 | -0.0706 | -0.824 | -- |
| 11 | 3/11/2009 12:59 | 1.59 | 1.47 | 39 | 29.6 | 1.59 | 39.7 | 0.000654 | 0.117 | -- |
| 12 | 3/12/2009 13:54 | 1.52 | 1.44 | 33 | 27.5 | 1.56 | 37.3 | -0.0447 | -0.358 | -- |
| 13 | 4/30/2009 15:02 | 1.53 | 1.42 | 34 | 26 | 1.54 | 35.6 | -0.011 | -0.0389 | -- |
| 14 | 5/27/2009 9:22 | 1.57 | 1.36 | 37 | 22.7 | 1.49 | 31.7 | 0.0759 | 0.719 | -- |
| 15 | 5/27/2009 14:18 | 1.56 | 1.48 | 36 | 30 | 1.6 | 40.2 | -0.0391 | -0.276 | -- |
| 16 | 8/7/2009 13:22 | 1.53 | 1.47 | 34 | 29.2 | 1.59 | 39.3 | -0.0539 | -0.53 | -- |
| 17 | 1/22/2010 9:24 | 2.02 | 1.91 | 105 | 81.4 | 1.96 | 93.9 | 0.0569 | 0.622 | -- |
| 18 | 2/1/2010 10:16 | 1.85 | 1.88 | 70 | 75.7 | 1.94 | 88.3 | -0.0924 | -0.939 | -- |
| 19 | 12/22/2010 12:30 | 1.58 | 1.34 | 38 | 22.1 | 1.48 | 31 | 0.0972 | 1.07 | -- |
| 20 | 3/23/2011 8:55 | 1.94 | 1.9 | 87 | 79.6 | 1.96 | 92.2 | -0.0167 | -0.117 | -- |
| 21 | 3/29/2011 12:38 | 1.9 | 1.91 | 80 | 80.5 | 1.96 | 93.1 | -0.0573 | -0.622 | -- |
| 22 | 1/27/2012 12:50 | 1.86 | 1.61 | 72 | 40.4 | 1.71 | 51.7 | 0.152 | 1.65 | -- |
| 23 | 2/24/2012 12:00 | 1.77 | 1.59 | 59 | 39 | 1.69 | 50.2 | 0.0785 | 0.824 | -- |
| 24 | 3/20/2012 15:05 | 1.59 | 1.37 | 39 | 23.3 | 1.5 | 32.4 | 0.0891 | 0.939 | -- |
| 25 | 4/3/2012 10:38 | 1.75 | 1.49 | 56 | 31.1 | 1.61 | 41.4 | 0.139 | 1.4 | -- |
| 26 | 6/26/2012 14:05 | 1.46 | 1.32 | 29 | 20.7 | 1.46 | 29.3 | 0.00411 | 0.196 | -- |
| 27 | 9/25/2012 12:12 | 1.18 | 0.785 | 15 | 6.1 | 1.01 | 10.4 | 0.169 | 2.08 | -- |
| 28 | 11/8/2012 10:53 | 1 | 0.634 | 10 | 4.3 | 0.878 | 7.69 | 0.122 | 1.22 | -- |
| 29 | 12/4/2012 12:48 | 1.9 | 1.79 | 79 | 61.8 | 1.86 | 74.3 | 0.0351 | 0.358 | -- |
| 30 | 12/6/2012 10:08 | 2.16 | 2.1 | 146 | 127 | 2.13 | 137 | 0.036 | 0.442 | -- |
| 31 | 12/6/2012 15:25 | 2.04 | 2.01 | 110 | 102 | 2.05 | 114 | -0.00771 | 0.0389 | -- |
| 32 | 6/26/2013 11:26 | 1.3 | 1.16 | 20 | 14.3 | 1.32 | 21.4 | -0.0205 | -0.196 | -- |

## Definitions

SSC: Suspended sediment concentration (SSC) in mg/l (80154)

Turb: Turbidity in FNU (63680)

## References


Domanski, M.M., Straub, T.D., and Landers, M.N., 2015, Surrogate Analysis and Index Developer (SAID) tool (version 1.0, September 2015): U.S. Geological Survey Open-File Report 2015–1177, 38 p., https://pubs.usgs.gov/of/2015/1177/ofr20151177.pdf.

Duan, Naihua, 1983, Smearing estimate – A nonparametric retransformation method: Journal of the American Statistical Association, Volume 78-383, 605-610 p.

Edwards, T.K. and Glysson, G.D., 1999, Field methods for measurement of fluvial sediment: U.S. Geological Survey Techniques of Water-Resources Investigations, book 3, chap. C2, 89 p., Available from: https://pubs.usgs.gov/twri/twri3-c2/pdf/TWRI_3-C2.pdf.

Helsel, D.R., and Hirsch, R.M., 2002, Statistical methods in water resources-Hydrologic analysis and interpretation: U.S. Geological Survey Techniques of Water-Resources investigations, book 4, chap. A3, 510 p.

Levesque, V.A., and Oberg, K.A., 2012, Computing discharge using the index velocity method: U.S. Geological Survey Techniques and Methods 3-A23, 148 p. Available from: https://pubs.usgs.gov/tm/3a23/.

R Core Team, 2018, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available from: https://www.R-project.org/.

Rasmussen, P.P., Gray J.R., Glysson G.D., Ziegler A.C., 2009, Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data: U.S. Geological Survey Techniques and Methods book 3, chap. C4, 53 p., Available from: https://pubs.usgs.gov/tm/tm3c4/pdf/TM3C4.pdf.

[USGS] U.S. Geological Survey, 2006, Collection of water samples (ver. 2.0): U.S. Geological Survey Techniques of Water-Resources Investigations. book 9, chap. A4. Available from: https://pubs.usgs.gov/twri/twri9a4/twri9a4_Chap4_v2.pdf.

[USGS] U.S. Geological Survey, 2014, Policy and guidelines for archival of surface-water, groundwater, and water-quality model applications: Office of Groundwater Technical Memorandum 2015.02, Office of Surface Water Technical Memorandum 2015.01, Office of Water Quality Technical Memorandum 2015.01, Available from: https://water.usgs.gov/admin/memo/SW/sw2015.01.pdf.

[USGS] U.S. Geological Survey, 2016, Policy and guidance for approval of surrogate regression models for computation of time series suspended sediment concentrations and loads: Office of Surface Water Technical Memorandum 2016.07, Office of Water Quality Technical Memorandum 2016.10, Available from: https://water.usgs.gov/admin/memo/QW/qw2016.10.pdf.

Wagner RJ, Boulger RW, Oblinger CJ, Smith BA, 2006, Guidelines and standard procedures for continuous water quality monitors: station operation, record computation, and data reporting: U.S. Geological Survey Techniques and Methods 1-D3, Available from: https://pubs.usgs.gov/tm/2006/tm1D3/pdf/TM1D3.pdf.